CLAIMS

What we claim is:

1   1.   An apparatus comprising an optimizing system that generates difference files
2   between an original byte stream and a new byte stream by identifying replacement
3   content in a first area of the new byte stream, wherein the replacement content includes a
4   group of bytes of the new byte streams identified as at least one of byte insertions and
5   byte replacements in an operation array, identifying content similarities between the
6   replacement content and at least one of the original byte stream and a second area of the
7   new byte stream, and encoding information of the content similarities to the difference
8   file.

1   2.   An apparatus for generating difference files, comprising:
2   means for receiving an original byte stream that is a segment of an original file
3   and a new byte stream that is a segment of a new file, wherein the new file includes an
4   updated version of the original file;
5   means for identifying replacement content of the new byte stream, wherein
6   replacement content includes a group of bytes of the new byte stream identified as at least
7   one of byte insertions and byte replacements in an operation array;
8   means for comparing the replacement content and portions of the new byte stream
9   and identifying a first set of content similarities resulting from the comparison;
10   means for comparing the replacement content and portions of the original byte
11 stream and identifying a second set of content similarities resulting from the comparison;
12 and
13   means for generating the difference file by encoding information of the first set
14 and second set of content similarities.

1   3.   The apparatus of claim 2, further comprising means for transferring the difference
2   file to a remote system that hosts a copy of the original file, the remote system updating
3   the hosted copy of the original file using the difference file.

1 4. A method for generating difference files, comprising:

2 receiving an original byte stream that is a segment of an original file and a new

3 byte stream that is a segment of a new file, wherein the new file includes an updated

4 version of the original file;

5 identifying replacement content of the new byte stream, wherein replacement

6 content includes a group of bytes of the new byte stream identified as at least one of byte

7 insertions and byte replacements in an operation array;

8 comparing the replacement content and portions of the new byte stream and

9 identifying a first set of content similarities resulting from the comparison;

10 comparing the replacement content and portions of the original byte stream and

11 identifying a second set of content similarities resulting from the comparison; and

12 generating the difference file by encoding information of the first set and second

13 set of content similarities.

1 5. The method of claim 4, wherein comparing the replacement content with portions

2 of the new byte stream further includes:

3 identifying at least one first largest common sub-string (LCS) of bytes in response

4 to the comparison; and

5 identifying the first set of content similarities between the replacement content

6 and the new byte stream using the first LCS.

1 6. The method of claim 4, wherein comparing the replacement content with portions

2 of the original byte stream further includes:

3 identifying at least one second LCS in response to the comparison; and

4 identifying the second set of content similarities between the replacement content

5 and the original byte stream using the second LCS.

1 7. The method of claim 4, further comprising:

2 encoding a first edit distance between the replacement content and portions of the

3 new byte stream; and

4   encoding a second edit distance between the replacement content and portions of

5   the original byte stream.

1   8.   The method of claim 7, wherein generating the difference file further includes

2   comparing information including at least one of a number of bytes used to encode the

3   first edit distance, a number of bytes used to encode the second edit distance, a number of

4   bytes of the replacement content, and a degree of content similarity.

1   9.   The method of claim 4, wherein generating the difference file by encoding

2   information of the first set of content similarities further comprises:

3   encoding a first edit distance between the replacement content and portions of the

4   new byte stream and encoding a second edit distance between the replacement content

5   and portions of the original byte stream;

6   comparing a number of bytes used to encode the first edit distance to a number of

7   bytes used to encode the second edit distance;

8   comparing the number of bytes used to encode the first edit distance to a length of

9   the replacement content and a degree of content similarity; and

10   encoding information of the first set of content similarities when the number of

11   bytes used to encode the first edit distance is less than or equal to the number of bytes

12   used to encode the second edit distance and a quantity formed by dividing the number of

13   bytes used to encode the first edit distance by the length of the replacement content is less

14   than the degree of content similarity.

1   10.   The method of claim 4, wherein generating the difference file by encoding

2   information of the second set of content similarities further comprises:

3   encoding a first edit distance between the replacement content and portions of the

4   new byte stream and encoding a second edit distance between the replacement content

5   and portions of the original byte stream;

6   comparing a number of bytes used to encode the second edit distance to a number

7   of bytes used to encode the first edit distance;

8   comparing the number of bytes used to encode the second edit distance to a length

9   of the replacement content and a degree of content similarity; and

10   encoding information of the second set of content similarities when the number of

11   bytes used to encode the second edit distance is less than or equal to the number of bytes

12   used to encode the first edit distance and a quantity formed by dividing the number of

13   bytes used to encode the second edit distance by the length of the replacement content is

14   less than the degree of content similarity.

1   11.   A system for updating electronic files of remote devices, comprising:

2   a first device including a file differencing engine that generates differences

3   between an original version and a new version of an electronic file by:

4   receiving an original byte stream that is a segment of the original version

5   and a new byte stream that is a segment of the new version;

6   identifying replacement content of the new byte stream, wherein

7   replacement content includes a group of bytes of the new byte stream identified as

8   at least one of byte insertions and byte replacements in an operation array;

9   comparing the replacement content and portions of the new byte stream

10   and identifying a first set of content similarities resulting from the comparison;

11   comparing the replacement content and portions of the original byte

12   stream and identifying a second set of content similarities resulting from the

13   comparison;

14   generating a difference file including encoded information of the first set

15   and second set of content similarities; and

16   a file updating engine hosted on the remote devices, the file updating engine

17   generating a copy of the new version using the difference file.

1   12.   The system of claim 11, wherein the remote device includes at least one of

2   cellular telephones, portable communication devices, personal digital assistants, personal

3   computers, and portable processor-based devices.

1   13.   A computer readable medium including executable instructions which, when

2   executed in a processing system, reduce a size of a difference file that includes coded

3   differences between an original version and a new version of an electronic file by:

4   identifying replacement content of a new byte stream that is a segment of the new

5   version, wherein replacement content includes a group of bytes of the new byte stream

6   identified as at least one of byte insertions and byte replacements in an operation array;

7   comparing the replacement content and portions of the new byte stream and

8   identifying a first set of content similarities resulting from the comparison;

9   comparing the replacement content and portions of an original byte stream and

10   identifying a second set of content similarities resulting from the comparison, wherein the

11   original byte stream is a segment of the original version; and

12   generating the difference file by encoding information of the first set and second

13   set of content similarities.